

Nicholaus Malone
IMT 531, Sam Oh
Final Project
March 12, 2009

Description of Collection

The collection I have chosen to describe in this project is the EBSCOhost Business Source Complete database. This database is dedicated to business topics such as information systems, management, strategy, economics, marketing, computer security, and more. The database contains articles from academic journals, trade publications, magazines, newspapers, books, company profiles, and more. The documents in the database can generally be accessed in both HTML and PDF format.

To manage and catalogue this collection of documents, EBSCO created an exhaustive amount of metadata for each document. In addition to the basic document identification data, such as author and publication, each document is associated with a number of controlled subject terms and author-supplied keywords. These aid users in searching for documents related to their topic of interest. Additionally, EBSCO has a thesaurus available to its users. This indicates that they have an extensive controlled vocabulary for use in cataloging their document collection. Unfortunately, they do not expose the structure of the thesaurus to the public. It is only available for alphabetical browsing or searching.

EBSCOhost Business Source Complete serves both professional and academic users, although the primary audience is likely to be academic users. The database is used to find documents related to specific topics or to find documents containing specific information. Although, users will occasionally know the exact article they want to find before accessing the site, generally, they will perform a search for their topic from the EBSCO interface. Since the database contains such a large quantity of documents and users typically have very specific needs, there are advanced searching filters that allow users a fine degree of control over the type of documents returned by a search. In order to fulfill these requirements, EBSCO must maintain extensive metadata records for their collection.

Description of Metadata Schema

In this project, I will create a metadata schema to describe this collection using elements borrowed from existing metadata standards. This project is not intended to be a complete implementation of any standards used in the schema. Rather, this project will use only those elements from standard schemas necessary to properly describe the collection and may exclude others.

To describe this collection, I have chosen to use a combination of descriptive metadata to describe the contents of documents and packaging metadata to maintain the digital files which comprise the collection. For the packaging and administrative metadata, I have chosen to implement a version of the Metadata Encoding and Transmission Standards (METS) published by the Library of Congress (LOC). I have chosen to extend METS with the lite version of the LOC's Metadata Object Description Schema (MODS) or MODS "Lite". MODS lite offers a simpler set of elements than the full version, yet it maintains enough descriptive power to describe the collection. Furthermore, all of the MODS lite elements correspond to the 15 core elements in the Dublin Core Metadata Initiative.

List of Elements

- doc
 - mets:dmdSec
 - mets:mdWrap
 - mets:xmlData
 - mods:mods
 - mods:titleInfo
 - mods:title
 - mods:name
 - mods:namePart
 - mods:typeOfResource
 - mods:dateIssued
 - mods:publisher
 - mods:abstract
 - mods:note
 - mods:subject
 - mods:topic
 - mets:fileSec
 - mets:fileGrp
 - mets:file
 - mets:FLocat
 - descData
 - industry
 - codes
 - numberPages
 - company
 - companyName
 - tickerSymbol

Data Dictionary

Ebsco Root & Custom Elements

doc

Namespace	http://www.ebscohost.com/bsc/
Definition	This element is the root container for a document in the database collection.
Optional	N
Repeatable	Y
Attributes	
Datatype	

doc\descData

Namespace	http://www.ebscohost.com/bsc/
Definition	Contains extra data describing the document used for searching
Optional	N
Repeatable	N
Attributes	
Datatype	

doc\descData\industry

Namespace	http://www.ebscohost.com/bsc/
Definition	A list of industry codes all provided by one code authority.
Optional	Y
Repeatable	Y
Attributes	<ul style="list-style-type: none"> • authority – the authority providing industry codes
Datatype	

doc\descData\industry\codes

Namespace	http://www.ebscohost.com/bsc/
Definition	An individual industry code describing the industry sectors to which this document pertains.
Optional	Y
Repeatable	Y
Attributes	
Datatype	NAICS controlled list of industry codes

doc\descData\numberPages

Namespace	http://www.ebscohost.com/bsc/
Definition	The number of pages in the document in its original format.
Optional	N
Repeatable	N
Attributes	
Datatype	integer

doc\descData\company

Namespace	http://www.ebscohost.com/bsc/
Definition	Contains data about the companies discussed in the article.
Optional	Y
Repeatable	N
Attributes	
Datatype	

doc\descData\company\companyName

Namespace	http://www.ebscohost.com/bsc/
Definition	The name of a company mentioned in the article.
Optional	Y
Repeatable	Y
Attributes	<ul style="list-style-type: none"> • type – official, alternative
Datatype	

doc\descData\company\tickerSymbol

Namespace	http://www.ebscohost.com/bsc/
Definition	The stock market symbol of a company mentioned in the article
Optional	Y
Repeatable	Y
Attributes	<ul style="list-style-type: none"> • exchange – NYSE, NASDAQ, LSE, HKEX A controlled list of stock exchange identifiers should be used
Datatype	A list of stock symbols controlled by the exchange being used.

METS Elements

The following elements and attributes are drawn from the METS metadata schema. These elements are also members of the root <doc> element.

dmdSec

Namespace	http://www.loc.gov/METS/
Definition	Container for descriptive metadata elements defined by the MODS schema.
Optional	N
Repeatable	N
Attributes	
Datatype	

dmdSec\mdWrap

Namespace	http://www.loc.gov/METS/
Definition	Wrapper for the descriptive metadata
Optional	N
Repeatable	N
Attributes	
Datatype	

dmdSec\mdWrap\xmlData

This element contains the “mods” element from the MODS namespace.

Namespace	http://www.loc.gov/METS/
Definition	Container for the MODS elements
Optional	N
Repeatable	N
Attributes	
Datatype	

fileSec

Namespace	http://www.loc.gov/METS/
Definition	Container for information about the physical files representing the document.
Optional	N
Repeatable	N
Attributes	ID
Datatype	

fileSec\fileGrp

Namespace	http://www.loc.gov/METS/
-----------	--------------------------

Definition	Represents a group of files associated with the document.
Optional	N
Repeatable	Y
Attributes	<ul style="list-style-type: none"> • USE – master, backup, public, etc...
Datatype	

fileSec\fileGrp\file

Namespace	http://www.loc.gov/METS/
Definition	Represents and provides access to a physical file.
Optional	N
Repeatable	Y
Attributes	<ul style="list-style-type: none"> • ID • MIMETYPE – can use IANA list • SIZE • CREATED
Datatype	

fileSec\fileGrp\file\FLocat

Namespace	http://www.loc.gov/METS/
Definition	Describes the physical location of the file.
Optional	N
Repeatable	N
Attributes	<ul style="list-style-type: none"> • LOCTYPE - location, ark, urn, purl, handle, doi, other • xlink:href
Datatype	empty

MODS ELEMENTS

The elements and attributes listed below are drawn from the MODSlite metadata schema. These elements are contained within the <dmdSec> element of the METS data.

Common Attributes

- lang – the ISO 639-2\b three letter language standard
- encoding – the data format used
- point – start or end

mods

Namespace	http://www.loc.gov/mods/v3
Definition	Container for all MODS elements
Optional	N
Repeatable	N
Attributes	
Datatype	

mods\titleInfo

Namespace	http://www.loc.gov/mods/v3
Definition	Wrapper for the titles of the document
Optional	N
Repeatable	N
Attributes	<ul style="list-style-type: none"> • type: abbreviated, translated, alternative, uniform • lang: iso 639-2\b
Datatype	

mods\titleInfo\title

Namespace	http://www.loc.gov/mods/v3
Definition	The title of the document
Optional	Y
Repeatable	Y
Attributes	
Datatype	

mods\name

Namespace	http://www.loc.gov/mods/v3
Definition	Represents the author of the document.
Optional	Y
Repeatable	Y
Attributes	
Datatype	

mods\name\namePart

Namespace	http://www.loc.gov/mods/v3
Definition	A section of the name
Optional	N
Repeatable	Y
Attributes	<ul style="list-style-type: none"> type: date, family, given, termOfAddress
Datatype	

mods\typeOfResource

Namespace	http://www.loc.gov/mods/v3
Definition	Describes the type of document
Optional	N
Repeatable	N
Attributes	
Datatype	Academic Journal, Periodical, Trade Publication, Newspaper, Book, Industry Profile, Country Report

mods\dateIssued

Namespace	http://www.loc.gov/mods/v3
Definition	The publication date of the document.
Optional	Y
Repeatable	N
Attributes	
Datatype	The W3CDTF date format

mods\publisher

Namespace	http://www.loc.gov/mods/v3
Definition	The title of the organization responsible for publishing the document
Optional	N
Repeatable	N
Attributes	

Datatype	
----------	--

mods\abstract

Namespace	http://www.loc.gov/mods/v3
Definition	A brief description of the document.
Optional	Y
Repeatable	N
Attributes	
Datatype	

mods\note

Namespace	http://www.loc.gov/mods/v3
Definition	Special notes pertaining to the document.
Optional	Y
Repeatable	Y
Attributes	
Datatype	

mods\subject

Namespace	http://www.loc.gov/mods/v3
Definition	A collection of topics pertaining to the document.
Optional	N
Repeatable	N
Attributes	
Datatype	

mods\subject\topic

Namespace	http://www.loc.gov/mods/v3
Definition	A specific topic to which the document pertains.
Optional	N
Repeatable	Y
Attributes	
Datatype	A controlled list of words defined in the EBSCO thesaurus